

AMENDMENTS TO THE CLAIMS

Please amend the claims as follows:

1. (Canceled).

2. (Canceled).

3. (Previously Presented) The method of claim 12, wherein comparing the seed article to at least one other related article is performed by a dynamic programming alignment algorithm to determine an alignment between the seed article and the related article.

4. (Previously Presented) The method of claim 12, further comprising determining a cluster of related articles from the related articles.

5. (Previously Presented) The method of claim 4, wherein determining the cluster of related articles is performed by:

using a dynamic programming alignment algorithm to compute edit distances
between the seed article and the related articles; and
choosing the cluster of related articles based on the edit distances.

6. (Original) The method of claim 4, wherein the identifying at least one information field within the seed article is performed by comparing the seed article to the cluster of articles.

7. (Previously Presented) The method of claim 12, wherein the information field corresponds to variable data.

8. (Previously Presented) The method of claim 12, wherein the articles are web pages.

9. (Original) The method of claim 8, wherein the related articles are web pages on a web site.

10. (Original) The method of claim 9, further comprising simplifying the content on a web page.

11. (Original) The method of claim 10, wherein simplifying the content includes preserving visible text, visible images, and visible paragraph and table formatting.

12. (Previously Presented) A method for information extraction, comprising:
accessing a plurality of related articles;
determining a seed article from the related articles;
identifying at least one information field within the seed article by comparing the seed article to at least one other related article;
creating a template based on the identified information field;
identifying a plurality of templates each comprising at least one information field;
comparing a source article to the templates to determine a closest template;
associating data from the source article with an information field from the closest template; and
extracting the associated data.

13. - 14. (Canceled).

15. (Previously Presented) A method of extracting data from a source article, comprising:
identifying a plurality of templates each comprising at least one information field;
comparing the source article to the templates to determine a closest template, wherein comparing the source article to the templates is performed by a dynamic programming alignment algorithm to compute an edit distance between the source article and the templates;
associating data from the source article with an information field from the closest template;
extracting the associated data; and
displaying the associated data.

16. - 18. (Canceled)

19. (Previously Presented) The computer program product of claim 23, wherein comparing the seed article to at least one other related article is performed by a dynamic programming alignment algorithm to determine an alignment between the seed article and the related article.

20. (Previously Presented) The computer program product of claim 23, further comprising computer program code for determining a cluster of related articles from the related articles.

21. (Previously Presented) The computer program product of claim 20, wherein determining a cluster of related articles is performed by:

using a dynamic programming alignment algorithm to compute edit distances
between the seed article and the related articles; and
choosing the cluster of related articles based on the edit distances.

22. (Previously Presented) The computer program product of claim 20, wherein the identifying at least one information field within the seed article is performed by comparing the seed article to the cluster of related articles.

23. (Currently Amended) A computer program product having a tangible computer-readable storage medium having computer-executable code encoded thereon for performing information extraction, the computer-executable code comprising code for:

~~a computer-readable medium; and~~
~~computer program code, encoded on the medium, for:~~
accessing a plurality of related articles;
determining a seed article from the related articles;
identifying at least one information field within the seed article by comparing the
seed article to at least one other related article;
creating a template based on the identified information field;

identifying a plurality of templates each comprising at least one information field;
comparing a source article to the templates to determine a closest template;
associating data from the source article with an information field from the closest
template; and
extracting the associated data.

24. (Previously Presented) The computer program product of claim 23, wherein
comparing the source article to the templates is performed by a dynamic programming alignment
algorithm to compute an edit distance between the source article and the templates.

25. (New) The computer program product of claim 23, further comprising code for:
displaying the associated data.

26. (New) The computer program product of claim 23, further comprising code for:
storing the associated data.